Journal of Physics: Conference Series 2017 vol.936 N1

Finding exact constants in a Markov model of Zipfs law generation

Bochkarev V., Lerner E., Nikiforov A., Pismenskiy A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© Published under licence by IOP Publishing Ltd. According to the classical Zipfs law, the word frequency is a power function of the word rank with an exponent -1. The objective of this work is to find multiplicative constant in a Markov model of word generation. Previously, the case of independent letters was mathematically strictly investigated in [Bochkarev V V and Lerner E Yu 2017 International Journal of Mathematics and Mathematical Sciences Article ID 914374]. Unfortunately, the methods used in this paper cannot be generalized in case of Markov chains. The search of the correct formulation of the Markov generalization of this results was performed using experiments with different ergodic matrices of transition probability P. Combinatory technique allowed taking into account all the words with probability of more than e -300 in case of 2 by 2 matrices. It was experimentally proved that the required constant in the limit is equal to the value reciprocal to conditional entropy of matrix row P with weights presenting the elements of the vector π of the stationary distribution of the Markov chain.

http://dx.doi.org/10.1088/1742-6596/936/1/012028

References

[1] Baayen R H 2001 Word Frequency Distributions (Dordrecht: Kluwer Academic Publishers)
[2] Bochkarev V V and Lerner E Yu 2012 Zipf and non-Zipf Laws for Homogeneous Markov Chain Preprint arXiv:1207.1872
[3] Bochkarev V V and Lerner E Yu 2012 Russian Mathematics 50 25
[4] Bochkarev V V and Lerner E Yu 2014 Electron. J. Linear Algebra 27 534
[5] Bochkarev V V and Lerner E Yu 2017 International Journal of Mathematics and Mathematical Sciences 2017 914374
[6] Conrad B and Mitzenmacher M 2004 IEEE Trans. Inform. Theory 50 1403
[7] Edwards R, Foxall E and Perkins T J 2012 Electron. J. Linear Algebra 23 966
[8] Jacquet Ph and Szpankowski W 1991 IEEE Trans. Information Theory 37 1470
[9] Sedgewick R and Flajolet P 1995 An Introduction to the Analysis of Algorithms (Boston Mass: Addison-Wesley)